DOCUMENT RESUME

ED 190 683 UD 020 671

AUTHOR Smith, Frederick
TITLE Testing: To What End?
PUB DATE [79]
NOTE 10p.: Tables may be marginally legible.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Administrative Policy: *Educational Objectives: Elementary Secondary Education: *Reading Tests: School Districts: *Testing Problems: *Test Interpretation: *Test Validity
IDENTIFIERS *New York (New York)

ABSTRACT

The 1978 New York City Test Report indicates that forty-three percent of those tested in that year in the City's public schools are reading at or above grade level. These results, however, when summarized in this way, are meaningless. The graph prepared for this article traces the reading levels of the school system and its 32 districts for 1974-78. The data reveal that the districts break into distinct levels of reading achievement with gaps between levels. Ten districts dwell near the norm, while four are above it and eighteen are below the norm. An equally important point is that the districts maintain a stability of rank over the five year span. Simply put, the annual reading test returns do not add to our knowledge. Citywide testing, as we've known it, must cease. The current system has led to misguided policy in the areas of teacher hiring and distribution of funds among districts. Objectives have been primarily administrative, rather than instructional. An entirely State-funded testing program has existed for some time and could be adapted within New York City to ensure continuity and comparability of results over time. A coordinated program with the State's testing procedures would, through tying measurement to instruction and skills development, better enable us to confront the problems of teaching children to read and maintaining educational standards.
(Author/GC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

MAY 2 9 1980

ED190683

TESTING: TO WHAT END?

by

Frederick Smith

Research Assistant

New York City Public Schools

UD020671

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Frederick Smith

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TESTING: TO WHAT END?

The 1978 New York City Test Report is out. The news is that 43% of those tested last year in our public schools are reading at or above grade level. But, once again, no comparisons are made with reading achievement in previous years. Instead, the Board of Education has issued its annual census of despair, a 118-page mound of numbers telling us far far too little.

And in the aftermath, we have just conducted the 1979 testing, to be followed in about a year by the 1979 Test Report—even though we can accurately predict the results of the next cycle of testing. As this analysis will show, we didn't have to give the 1979 test, and now that we have, we don't have to wait another year for the results. Giving the test each year has become an end in itself.

The ritual is unacceptable given the stakes. To begin with, there is a large investment of time and money in the testing program. In 1978, for example, 507,000 pupils in grades 2 to 9 and 30,000 teachers spent an entire school day engaged in testing. While test materials and scoring cost $420,000, the manpower involved in administering the tests came to $2.5 million. So the rites of Spring '79 have consumed about another $3.0 million. In addition, there is the price paid every year by teachers and students alike, made anxious to do well. All of this sacrificed on the altar of testing, year after year, for no comparisons, no answers.

Surely the results demand interpretation in ways that make educational sense. The graph prepared for this article traces the reading levels of the school system and its 32 community school districts in terms of the percentage of pupils tested who were found to read at or above the norm in each of the last five years. (As a point of reference, being at the norm or reading on grade level is represented by the horizontal fifty-percent line.)

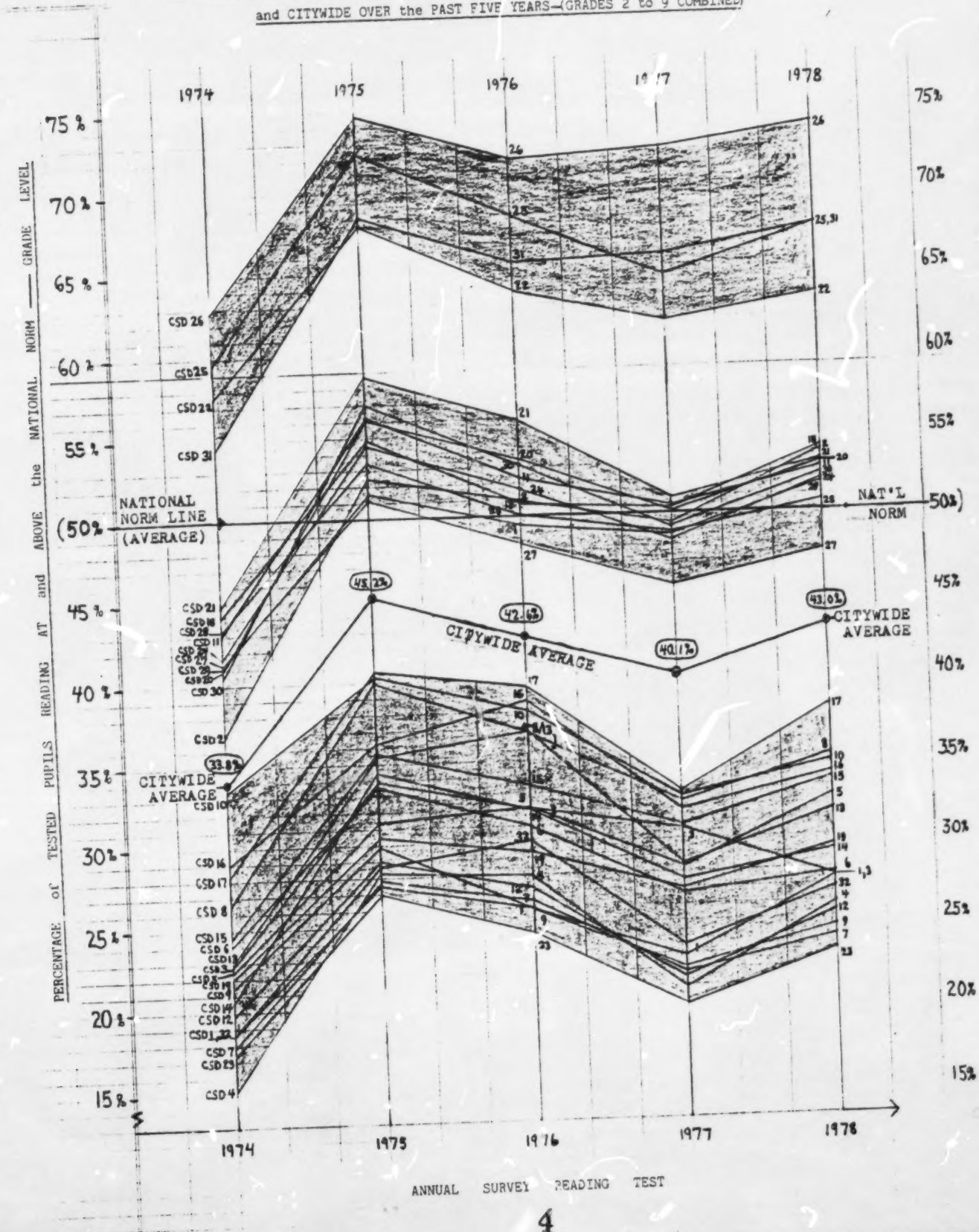
Now we can see why a statement that 43% of 1978's students are at or above the norm is less than informative. The percentage reflects no reality and describes no district. In 1978, as in past years, it falls into the divide that exists between the higher and lower scoring districts. It is this atypical figure, however, that has been taken to characterize reading achievement for the whole system.

The graph reveals that the districts break into distinct levels of reading achievement with gaps between levels. It tells a tale of separate systems. Ten districts dwell near the norm. Four are above it. Eighteen are below it. No comparisons? We should be developing all of our data in order to identify and meet the learning needs of each level.

An equally important, if less obvious point is that the districts maintain a stability of rank over the five-year span. There is very little shifting in their relative positions. Generally, as the figure illustrates, when the districts go up, they all do. When one declines, they decline together. And the very reason given in the 1978 Test Report for not making comparisons—different tests were used from one year to the next—makes the graphic depiction all the more forceful. In short,

PUBLIC SCHOOL READING ACHIEVEMENT LEVELS AMONG the 32 COMMUNITY SCHOOL DISTRICTS (CSD)

and CITYWIDE OVER the PAST FIVE YEARS—(GRADES 2 to 9 COMBINED)



there seems to be a holding pattern in New York City reading performance that overrides all other factors.

This leads to the realization that, in the present testing program, the perennial debate over whether our reading levels have gone "up" or "down" is of little relevance, if not chimerical. From the time-lapse perspective of the graph, the ups and downs reflect little more than the use of different tests. The point is accentuated in the otherwise inexplicably sharp upswing between 1974 and 1975. It is not flippant to suggest that from year to year the New York City test trends really tell more about the test series that happened to have been used than about the specific state of reading in our schools. As often as the tests change, the more reading achievement seems to stay the same.

The strength of the observed holding pattern is indicated by a few statistics. So nearly perfect is the correlation between the 1977 and 1978 inter-district ranks that 1978's overall standings (all grades combined) prove to be 96% predictable from the 1977 results—just as 1977's were also 96% predictable from the 1976 results. If reading achievement is taken as height, then we can project, with high probability, that the districts will remain in their size places.

Simply put, the annual reading test returns do not add to our knowledge. Subanalyses of the data support this generalization. For instance, in grade 2 alone, the '77 and '78 line-ups have 95% in common, the same as the year before. Within grade 5, 1977 and 1978 have a 92% overlap, and the correspondence in grade 8 is 91%. With so strong a degree of mutuality in the districts from one year to the next, we need not conduct a yearly test survey to know its likely outcome. For, on this level of analysis, the testing program supplies information that continues to be redundant.

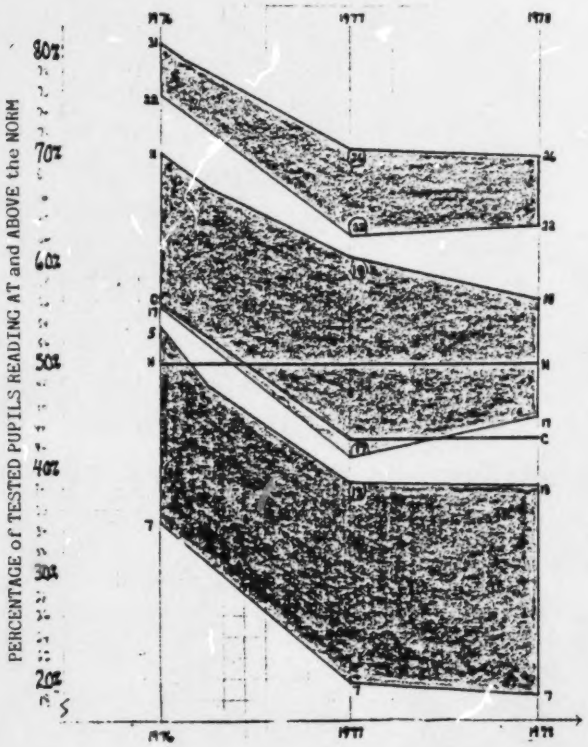
The graph and statistics not only trace deficiencies in our testing. They are starkly disquieting about the educational processes being measured as can be further discerned by plotting the results of the last three years for particular grades. Graphs have been drawn for grades 2, 5 and 8; these are derived from and underlie the composite results. They delineate the prevailing theme of stratification among the districts into levels of achievement. Although there is a slight shifting within each band of districts, no district succeeds in crossing the divide. Perhaps the most distressing deduction is that the differences are manifest as early as grade 2, and these are differences which, up to now, the districts have been unable to overcome. To imply, as these findings would seem to, that we cannot change the order of things jolts a basic premise of education.

One other observation affirms that the reading crisis is deep as well as wide. No one in the ranks can be smug because even in the top reading district almost 375 children are two or more years below the norm—enough to fill an elementary school building. Another district has over twelve times that number in its classrooms, and the City has 68,000 pupils with such extreme reading deficits.

While graphs and correlations are not explanations, they challenge us to discover what may be operating to cause X and Y to go together so consistently and, if possible, to do something about it. It must be made explicitly clear that the data presented here do NOT establish: that decentralization is a failure; that school districts actually

RELATIVE LEVELS of READING ACHIEVEMENT AMONG
NEW YORK CITY PUBLIC SCHOOL DISTRICTS OVER the LAST THREE YEARS
(GRADES 2, 5 and 8)

GRADE 2



Explanation:

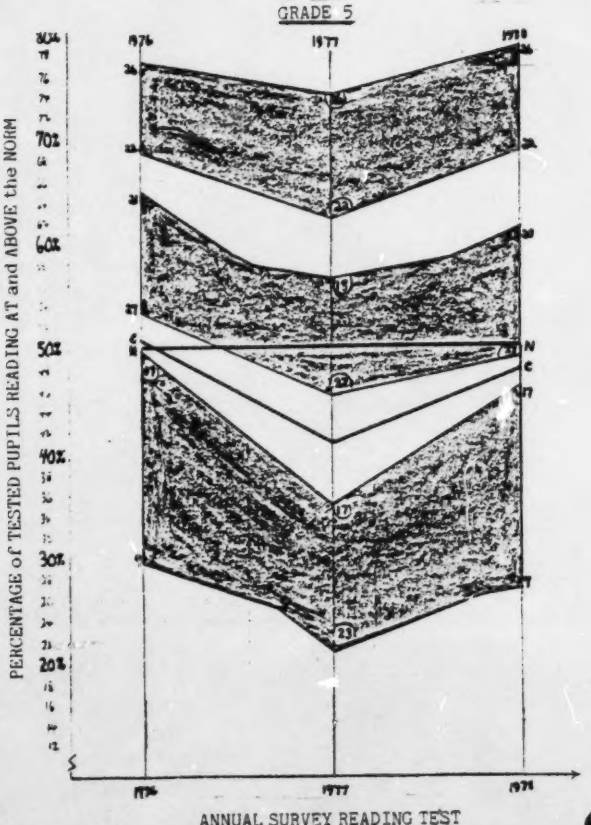
N = National Norm

C = Citywide Norm

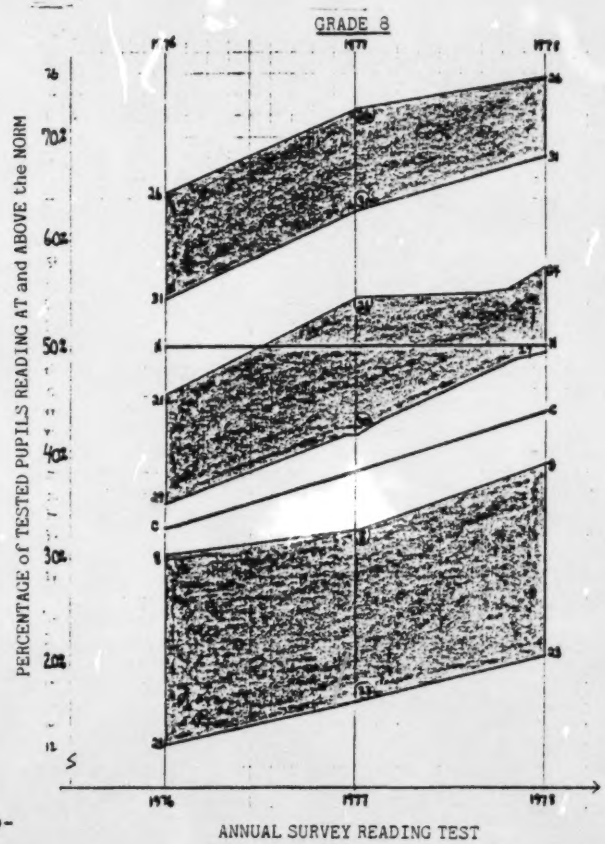
Numbers along axes indicate school districts bounding separate bands of districts.

The graphs for grades 2, 5 and 8 are derived from the composite on page 2.

GRADE 5



GRADE 8



widen disparities; that principals and teachers aren't worth what they're paid; that there should be a moratorium on all testing; or that minority group pupils are inferior learners; that schools cannot be expected to cope with the larger problems that come through the classroom door.

Further, with the focus on the district as a unit of analysis, the data do not deny the existence of exemplary schools or programs. To the contrary. This treatment of data suggests a methodology that should be applied to schools as well as districts in order to aid in isolating and studying instances of school success.

What the data discussed in this article do demonstrate is that the present citywide testing program has been a dead end. We must believe that the public schools may bring about positive changes in the lives of their students. So if the districts, schools and teachers are to make a difference, the unthinking type of testing and reporting that has gone on to date certainly doesn't show the way. A first step, then, is to re-examine our testing.

* * * *

One conclusion to be drawn from the empty testing record is that the citywide testing, as we've known it, must cease. No useful information would be lost because 1979's results could have been forecast even before the test was given. Abolishing the present program would save the City half a million dollars a year plus another $2.5 million in good money thrown after it. And everyone would be spared round after round of the same insightless routine.

Legislation is the key. The 1969 Decentralization Act requires that all children in the districts be tested annually in reading. The mandate is to rank the district schools in order of reading achievement based on the percentage of pupils at and above the norm. As intended, such use of the tests has allowed those local districts with schools low in reading level more flexibility in the selection of their teachers.

This has proven to be a misguided policy because a two-track hiring system has developed. The districts poorest in basic reading skills have been enabled to apply what many observers agree are the relatively loose selection standards of the National Teacher Examination. It is this alternate method of teacher selection that has come under increasing criticism by those who feel the practice has created a double standard, diluted the quality of the teaching pool in the districts of greatest need, and further harmed our most vulnerable readers. In fact, although the 18 districts which have remained in the lowest stratum of reading achievement contain 50% of the decentralized schools, they account for 87% of all schools eligible for dual hiring. A fair case can be made that the legislation defining the test program has unwittingly worked against improved educational performance.

The citywide reading test has also been set to other organizational purposes. Reading scores figure in the school system's delicate policies regarding pupil promotion as well as distribution of funds among the districts. Here, too, test scores have been weighed into murky formulas and the effects are being denounced.

The singular fact is that while the reading survey has come to serve

administrative purposes, the most important use of a reading test has been neglected: helping teachers help each pupil acquire reading skills and fostering the learning process. Just ask the teachers whether they are satisfied with the educational yield of the testing program. This, in the era of accountability.

With the over-emphasis on testing to meet administrative rather than instructional objectives, it's as though public officials have let tests become their substitute for responsible thinking. In deciding to let test scores "determine" how teacher hiring takes place, which pupils get promoted, and where the money goes, legislators and school officials distance themselves from three hot political issues. Then, with so much riding on it, the testing must go on, autonomous, relentless.

A few observations are offered as a way out. First, it is clear that the testing program has centered on matters internal to the school system—staffing patterns, promotional standards, and apportionment of funds. As such, there has been no administrative (or legislative) imperative here for giving a nationally normed test. The scores are used to make choices within and among the 32 local districts. A national norm adds nothing to these kinds of intra-organizational judgments.

Second, an entirely state-funded testing program has existed for some time and could have been adapted to accomplish the same institutional ends. The annual New York State Pupil Evaluation Program measures the reading achievement of all youngsters in grades 3 and 6 and the basic competency of students in grade 9. The local results are reported each year. These scores could easily be used to rank school performance and to guide allocations. What's more, this parallel program functions well without national norms. It does so because the State constructs its own tests, insuring continuity and comparability of results over time.

New York City's program does not have these advantages mainly because the Board of Education has chosen to purchase different tests from commercial publishers. Turning to "outside experts" to supply our testing sealed the Board's abdication in the matter. Private interests took over. In the transaction the City, representing the largest public school testing market in the country, was sold a program that lacks comparability of results and makes up for this by providing norms we don't need.

One can only surmise that school authorities were induced to believe that the act of measuring ourselves against national norms each year was necessary to the preservation of educational standards. That the Board members still cling to their gross survey of reading, confusing norms with standards, marks the degree to which they are acting without perspective.

An effective testing program must be more than a once-a-year imposition; it must have direct meaning for the classroom teacher. The Board of Education should view reading as a set of specific skills to be mastered. Defining the reading process this way makes mastery of each underlying skill an educational objective, with testing designed for the discrete objectives. The approach is really a refined type of diagnostic testing, enabling teachers to find out what the individual pupil can do and to identify those particular reading skills the pupil has not yet attained. Instruction is tailored accordingly as pupils progressively develop the requisites of reading comprehension. The logic of a sound program would

extend to the design and evaluation of reading curricula and to the improvement of instruction. The strategy is not new. Its systematic application in the public school population would be.

The one sign that things could be changing has been the action of the new Schools Chancellor. He has called for a diagnostic testing program to begin this Fall, covering all pupils in grades 2 through 8. He has also given primacy to strengthening education in the earliest school years. Both of these efforts are consistent with this article's findings.

In light of the Chancellor's perception, however, and a concern with spending money that will pay off in the classroom, it is surprising that he is still going along with the traditional Spring reading test. If anything, it would seem that having a diagnostic test in place in grades 2 to 8 solidifies all the points against continuation of the survey test in grades 2 to 9, setting up the conditions in which a rational testing program and state legislation could be brought into harmony.

In a coordinated program, the State's test would find its correct place next to our diagnostic approach. Additionally, whatever value national norms hold for educators could be derived from a nationally standardized test given to a five-percent sample of pupils, and we would realize a huge cost-benefit. We could then afford to spend more time and care testing the small representative sample in greater depth, and this thorough assessment of reading ability would tell us more, with less disruption, about the status of reading in the public schools than we could ever learn by going through the motions of testing everyone superficially. For this plan to have full utility, however, the Board needs to set up its own test research and development unit and take charge of testing lest another publisher once again sell us a buy now, pay later test package. We have the resources and the obligation to deliver testing that is interpretable.

Tying measurement to instruction, with emphasis on skills development and objectives, would put us in a better position to confront our most intractable problems which are teaching children to read, maintenance of educational standards, and the related complexities of pupil promotion. The inability to solve these problems finds its ultimate expression in the fear some educators have of the impending use of a Basic Competency Test as a high school graduation requirement—a moment of truth for the educators and a final grade for those students groping for the finish line after twelve years of public education.

In the long run, one might even hope that testing would come to be valued rather than feared. This could be the case if the Board reconsidered its test priorities and methods. Then, the testing program would become more than a deathwatch, and the Test Report would be more than a body count.

NOTE:

On May 14, 1979 Chancellor Macchiarola directed that the scores of all 1979 achievement tests given in March be withheld from use pending full investigation of possible technical flaws in the exam. Pending outcome of the review he directed that payment be withheld from the publisher, Harcourt Brace Jovanovich. The price of the test in question, Form K of the Metropolitan Achievement Test was $499,924. The Chancellor indicated that if the review caused the 1979 test to be invalidated, another achievement test would be given.

On June 6, 1979 all students in grades 2 through 9 took a second citywide reading test reportedly because of flawed procedures followed by the publisher of the first exam. The second test was published by McGraw-Hill Inc.

Frederick Smith has been a licensed researcher with the New York City public schools since 1969. While appointed to the Office of Educational Evaluation he worked in the area of achievement testing.